HIGH SPEED DATA PROCESSING AT THE WESTERBORK SYNTHESIS RADIO TELESCOPE

Mark Bentum

ASTRON
Oude Hoogeveensedijk 4, 7991 PD Dwingeloo, The Netherlands
bentum@astron.nl

ABSTRACT

The new Westerbork Synthesis Array has a correlator capacity of 256.000 complex channels. The number of channels must be distributed over the interferometers, number of lags and polarizations. The raw output of the correlator has a data rate of more than 20 Gbit/s. At this moment this data is integrated over 10 seconds before further processing is done. Therefore we can cope with the data rates for normal operation. For specific astronomical projects, a faster readout is required. A new project is started to read out the raw data from the backend. In this paper a description of the current setup is given as well as a glance to the future.

INTRODUCTION

Since the end of 2001 there has been a brand new Westerbork Synthesis Radio Telescope (WSRT) standing in the northern part of The Netherlands. The upgrade is finished and the new frontend and backend capabilities are enormous. This paper focuses on the current data rates in the backend system. We will start with an inventory of the current data flow of the (digital) signals at the WSRT, from the first digitisation to the final product of the WSRT, a Measurement Set (MS) of each observation. After each observation a number of actions are performed with the data. After the discussion of the data handling a number of future additions will be presented, such as the fast data readout from the correlator and real-time VLBI.

The WSRT consists of 14 telescopes, each with a diameter of 25 m. Ten of the dishes have a fixed location, while two at the eastern end of the array can be moved on rails. 1.4 km to the east is a second pair of movable dishes on rails. In each of the telescopes a multi-frequency frontend is mounted, with observing capabilities in the spectrum from 250 MHz up to 8.7 GHz. In the MFFE's the sky signals are mixed to intermediate frequency with a central frequency of 100 MHz and a bandwidth of maximum 160 MHz. The MFFE's have dual polarization receivers, therefore a total of 28 signals enter the building. After equalizing (cable length is about 1 km) the signals are handled by the backend. The first step in the backend is the division of the 160 MHz wide analog signal in eight subbands of (maximum) 20 MHz. This is done in the Intermediate to Video Conversion (IVC) system. The fringe correction is also performed in the IVC. After the IVC the signal are digitised in the Analog-to-Digital Conversion (ADC) systems. Delay corrections are performed in the ADC as well. After the ADC the correlation is done in the Correlator (COR) systems.

THE NEW BACKEND OF THE WSRT

The new backend, the DZB, broadens the bandwidth from 80 to 160 MHz divided into 8 sub-bands of 20 MHz. The central frequency of these sub-bands can be tuned to any desired frequency within the 160 MHz band. In Fig. 1 an impression of the current backend hardware is given.

The sensitivity is improved further by using 2-bit correlation and by correlating all 91 interferometers (in case of 14 telescopes) in 4 polarizations. For line observations there will be 64 complex channels for each of the 20 MHz sub-bands yielding a total of 512 channels over the 160 MHz bandwidth range for each interferometer-polarization. The channel separation is then 0.3125 MHz. In addition to all 91 cross-correlations the DZB also provides all 14 auto-correlations (for all polarizations). It accepts in fact a total of 16 inputs corresponding to 120 interferometers and 16 autocorrelations. The two additional inputs may come from the tied-array adder outputs or from two more antennas. For line observations that require a bandwidth of maximally 20 MHz the number of complex channels over that band will be equal to 512 for all 4 polarizations and equal to 1024 when only 2 polarizations are needed.

Fig. 1. The current backend hardware at the WSRT. Successively we see the IVC (Intermediate to Video Convertor system), one of the ADCs (Analog to Digital Convertor) and part of the correlator.

DATAFLOW

The IF signals from the equalizers are mixed to video (0-20 MHz) in the IVC unit. The 160 MHz band is divided in 8 bands. Therefore a total of 256 video-channels (16 telescopes * 2 polarizations * 8 bands) are send to the ADCs. The overall data rate coming from the analog-to-digital converter subsystem is equal:

2 (bits) *256 (channels) * 40 MHz (sample rate) = **20.480 Gbit/s**

Currently these signals are sent, via a data distributing unit (DDU), to the correlator. The DDU can be used for selecting the data to be send to all the correlator crates. The total capacity of the correlator is 256.000 correlator lags. Every BOCF (Beat of correlator frame) the correlations of all the interferometer arms are performed in the correlator. The BOCF is 1024 frames per 10 seconds. At this moment the data is integrated to basic intervals of 10 seconds. (Note that in the future we will access the raw data in the BOCF of 10ms!). The next step in the process is the transformation of the correlated lag data into the frequency domain. Most of the astronomical data processing software works on data in the frequency domain. This task is done in the data acquisition workstation (a HP 9000 – C3600 workstation). After normalization, some corrections and, if needed, further integration, the data is stored in a Measurement Set (MS) data format. In Fig.2 a schematic is presented of the current processing pipeline of the WSRT.

The size of the MS depends on the following parameters:

- N_i = Number of interferometers (= 120 cross and 16 auto correlations)
- N_b = Number of bands (maximum of 8)
- N_f = Number of frequency points (e.g. 128)
- N_p = Number of polarizations (either 1,2, or 4)
- N_t = Number of telescopes
- N_s = Number of samples (in case of 10 seconds integration time and 12 hours of observation, N_s equals 6*60*12 = 4320)

Besides the data, also a certain amount of overhead is present in the MS. In (1) the equation of the size of a MS is given. This equation is empirically found [1].

$$\text{Size MS (Mbit)} = 662765 + (110 + N_f*(8.125 * N_p + 4))*N_i*N_b*N_s + 377*N_t*N_b*N_s \qquad (1)$$

In case of maximum use of the correlator, the total size of the MS (in case of 10 seconds integration time) is more than 12 GByte, meaning a dataflow of 1 GByte/hour.

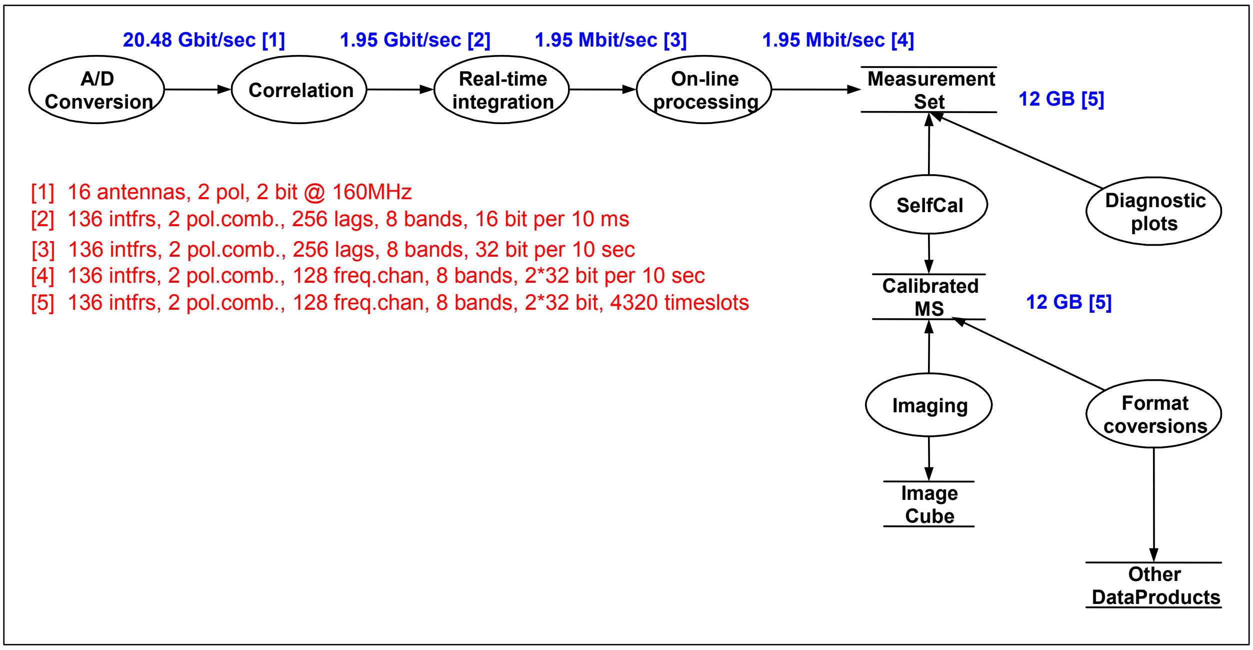

Fig.2. The processing pipeline of the data in the backend including the data rates.

DATA HANDLING

The MS with a maximum size of 1 GByte per hour is temporarily stored on the data acquisition computer. After the observation is finished, a couple of actions are started:

- Safeguard of the data by copying the MS to the archive server.
- Copy the MS to the inspection server from the archive server, to be ready for further processing.
- Extract information from the MS for supporting databases (e.g. The online available table with limited information of all observations done since 1970).
- Auto-analysis of the data for faults.
- Extraction of system parameters to feed them back into the system (e.g. phase and delay offsets, pointing parameters of the telescopes, attenuator settings).
- Production of inspection plots (e.g. passbands, Tsyst, Tnoise, gain, total power).
- Creation of administrative information available for the online (www) documentation.
- Conversions of the MS to other formats (UVFITS, MSFITS and Newstar scanfiles).

Besides these operations we will also archive every observation. Currently this is done on tape and CDROM's. Export of the data to the Primary Investigator (PI) will be done within two weeks after the observation. Doing this, storing capacity of 3 formats (MS, SCN, UVFITS file format) is required for at least 2 weeks. This is maximum 1 TByte of information. At the WSRT two RAID systems are present, each capable of storing 0.5 TByte of data.

Currently, we are looking for the possibility to archive the data on large disk arrays. As technology progresses, disk arrays will become larger and/or cheaper per GByte. The advantages of such a system are:

- Permanent online available data.
- Secure storage because of the RAID standards and possibility of constant monitoring of the status of the hardware (and also on the status of the data).
- Easily scalable in capacity.
- Technology independent.
- Minimum user interaction required.
- Possibility for an online virtual observatory.

To establish good control of the dataflow, without interference in the process of capturing data, three separate networks are implemented. The first network (called the real-time network) consists of all workstations dealing with the control and capturing of the data (like the real-time VME computers of the ADCs, DDU and the COR). The second network is the so-called online network. In this domain all the computers required for the operation of the telescope are connected (like the telescope control workstation and the central TMS (Telescope Management System) workstation). The third network is the offline network. This is the *astron.nl* domain and connects all offline computers in Westerbork (like desktop computers of the employees, but also the offline data processing workstations). Only authorized persons can access the online and real-time domains, to ensure continuous data acquisition. This turned out to be very successful.

FUTURE PLANS

Several projects are or will be started regarding the data and dataflow of the WSRT.

Reduction pipeline

Data rates for normal operation are under control. We can observe continuously with the highest data rates, including full backup, checking of the data and export within two weeks. Still, the data sizes are quite large (up to 12 GByte) and not always necessary for the astronomical purpose. If the PI is only interested in the continuum flux of the observed radio source, not all the line data is needed and a considerable reduction can be performed. This meaningful reduction of the data is called the reduction pipeline. At the WSRT we decided to run the reduction pipeline on observations that are already observed. So we archive the full observation and export the reduced data. Doing this, it is possible to use the archived data for other purposes or redo the reduction in case something went wrong. Another possible reduction algorithm is to eliminate RFI artefacts in the data.

Fast readout of the backend

The DZB in principle correlates the data in 10 ms intervals. Due to the huge amount of data, integration in 10 seconds intervals is currently done. The Post Correlator Integrator [2] is a device that will enable fast readout of the DZB. The current hardware of the DZB is not able to support this huge amount of data and must therefore be equipped with new hardware. Having the data available, a couple of things can be done. One example is applying a real-time RFI reduction algorithm on the data and after completion, feed it back in the data chain. Another use is to reduce the basic integration interval (e.g. 1 second). The 10 ms output has a huge impact on the data rate. The complete data flow, for 8 bands of 20 MHz bandwidth, is 244.15 MBytes/s.

Real-time VLBI

Another project is enabling real-time VLBI. Currently the data is stored on magnetic tapes. After completion, the tapes are shipped to the central correlator (for EVN in Dwingeloo) where the data is correlated. The future plans for VLBI are to connect the VLBI stations directly with the correlator centre (via fiber connections). If the complete band of 160 MHz will be used with 2 polarizations and 2 bits Analog-to-digital conversion, a data rate of 1.2 Gbit/s will be needed.

CONCLUSIONS

In this paper an overview is given of the current dataflow, data rates and data processing of the ‘new’ WSRT. The amount of data is huge and will even be more in the future if the fast readout option is available. On the other side, we are working on automatic reduction pipeline algorithms for considerable reduction of the data.

ACKNOWLEDGEMENTS

The work presented in this paper could be made thanks to the effort of many people in ASTRON. Especially the work of Albert Bos, Hans van Someren-Greve, Teun Grit and Arno Schoenmakers is acknowledged.

REFERENCES

[1] H. van Someren Greve, “The size of the Measurement set”, internal report ASTRON
[2] H. Verkouter, “Design specification for the Post-Correlator Integrator in the EVN MkIV correlator”, internal design document JIVE.